Cost function vs Loss function.

- when we calculate $J(\beta_0, \beta_1)$ for each individual datapoints separately it's called cost function vs when we calculate $J(\beta_0, \beta_1)$ for all n datapoints at once it's loss function.

$$\boxed{\text{Cost function: } J(\beta_0, \beta_1) \quad \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

$$\boxed{\text{Loss function: } J(\beta_0, \beta_1) \quad (y_i - \hat{y}_i)^2}$$

y_i : Actual value of variable

$\hat{y}_i$: Predicted value of variable

n : Total no.s of datapoints.

From convergence Algorithm we have:

$$\beta_i = \beta_i - \alpha \left[\frac{\partial}{\partial \beta_i} (J(\beta_i)) \right]$$

also, from cost function

$$J(\beta_i) = J(\beta_0, \beta_1)$$

β_0 : value of y if x is zero [Intercept]

β_1 : Slope

$$J(\beta_0, \beta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

For $J = 0$, above equation become:

$$J(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

For $J = 1$,

$$J(\beta_1) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

Now finding the derevative using convergence algorithm.

$$\frac{\partial}{\partial \beta_j} J(\beta_j)$$

Case I : $J = 0 \quad (J(\beta_0))$

$$= \frac{\partial}{\partial \beta_0} J(\beta_0)$$

$$= \frac{\partial}{\partial \beta_0} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2$$

$$= \frac{2}{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]$$

Case II : $J = 1 \quad (J(\beta_1))$

$$= \frac{\partial}{\partial \beta_1} J(\beta_1)$$

$$= \frac{\partial}{\partial \beta_1} \frac{1}{n} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2$$

$$= \frac{2}{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right] \times x_i$$

Combining above derivatives for $J = 0$ and $J = 1$ for the repeat until convergence.

we get :

$$\boxed{\left\{ \begin{aligned} \beta_0 &= \beta_0 - \alpha\left[\frac{2}{n}\sum_{i=1}^{n}\left(y_i - (\beta_0(x)_1)\right)\right] \\ \beta_1 &= \beta_1 - \alpha\left[\frac{2}{n}\sum_{i=1}^{n}\left(y_i - \beta_1(x_i)\right)\right]x_i \end{aligned} \right\}}$$

$\beta_0(x_1) = \beta_0 + \beta_1 x_1$

$\beta_1(x_i) = \beta_0 + \beta_1 x_i$

x_i = x value of i^{th} term

n = Total no.s of datapoint.